

Random intersection graph process

Mindaugas Bloznelis* and Michał Karoński†

Abstract

We introduce a random intersection graph process aimed at modeling sparse evolving affiliation networks that admit tunable (power law) degree distribution and assortativity and clustering coefficients. We show the asymptotic degree distribution and provide explicit asymptotic formulas for assortativity and clustering coefficients.

keywords: random graph process, random intersection graph, degree distribution, power law, clustering, assortativity

2000 Mathematics Subject Classifications: 05C80, 05C07, 05C82

1 Introduction

Given non-negative weights $x = \{x_i\}_{i \geq 1}$ and $y = \{y_j\}_{j \geq 1}$, and a nondecreasing positive sequence $\{\tau(t)\}_{t \geq 1}$, satisfying $\lim_{t \rightarrow +\infty} \tau(t) = +\infty$, let $H_{x,y}$ be the random bipartite graph with bipartition $V = \{v_1, v_2, \dots\}$ and $W = \{w_1, w_2, \dots\}$, where edges $\{w_i, v_j\}$ are inserted independently and with probabilities

$$p_{ij} = \min\left\{1, \frac{x_i y_j}{\sqrt{i j}}\right\} \mathbb{I}_{\{a\tau(j) \leq i \leq b\tau(j)\}}. \quad (1)$$

Here $b > a > 0$ are fixed numbers. $H_{x,y}$ defines the random intersection graph $G_{x,y}$ on the vertex set V such that any $u, v \in V$ are declared adjacent (denoted $u \sim v$) whenever they have a common neighbor in $H_{x,y}$.

Consider, for example, a library where a new item w_i is acquired at time i , and where a new user v_j is registered at time j . User v_j picks at random items from a "contemporary literature collection" $\{w_i : a\tau(j) \leq i \leq b\tau(j)\}$ relevant to time j (the interval $\{i : a\tau(j) \leq i \leq b\tau(j)\}$ can also be considered as the lifetime of the user v_j). Every actor v_j and every item w_i is assigned weight y_j and x_i respectively. These weights model the activity of actors and attractiveness of literature items. Now, assume that up to time t the library has acquired items $\{w_1, \dots, w_{\tau_*(t)}\} =: W_{\tau_*(t)}$, where $\tau_* : \mathbb{N} \rightarrow \mathbb{N}$ is a given nondecreasing function satisfying $\lim_{t \rightarrow +\infty} \tau_*(t) = +\infty$. The subgraph $H_{x,y}(t)$ of $H_{x,y}$ induced by the bipartition $V_t = \{v_1, v_2, \dots, v_t\}$ and $W_{\tau_*(t)}$ defines the random intersection graph $G_{x,y}(t)$ on the vertex set V_t : vertices $u, v \in V_t$ are declared adjacent whenever they have a common neighbor in $H_{x,y}(t)$. The graph $H_{x,y}(t)$ represents a snapshot taken at time t of the "library" records, while the graph $H_{x,y}$ shows the complete history of the "library". Graphs $G_{x,y}(t)$ and $G_{x,y}$ represent adjacency relations (between users) observed up to time t and during the whole lifetime of the "library", respectively. Assuming, in addition, that x and y are realized values of iid sequences $X = \{X_i\}_{i \geq 1}$ and $Y = \{Y_j\}_{j \geq 1}$ we obtain the

*Faculty of Mathematics and Informatics, Vilnius University, 03225 Vilnius, Lithuania

†Faculty of Mathematics and Computer science, Adam Mickiewicz University, 60769 Poznań, Poland

random graph $G_{X,Y}$ and the random graph process $\{G_{X,Y}(t)\}_{t \geq 1}$. The parameters of such a network model are the probability distributions of X_1, Y_1 , the functions τ, τ^* and the cutt-offs $a < b$.

Random intersection graph $G_{X,Y}$ is aimed at modeling sparse evolving affiliation networks that admit a power law degree distribution and non-vanishing *clustering* and assortativity coefficients. We first observe that choosing inhomogeneous weight sequences x and y one typically obtains an inhomogeneous degree sequence of the graph $G_{x,y}$: vertices with larger weights attract larger numbers of neighbours. Consequently, in the case where the probability distributions of X_1 and Y_1 have heavy tails, we may expect to obtain a heavy tailed (asymptotic) degree distribution in the random graph $G_{X,Y}$. Secondly, we observe that if the set $W(t)$ of items selected by a user v_t is (stochastically) bounded and the lifetimes of two neighbours of v_t , say v_s and v_u , intersect, then with a non-vanishing probability v_s and v_u share an item from $W(t)$. Consequently, the conditional probability $\alpha_{t|su} = \mathbf{P}(v_s \sim v_u | v_s \sim v_t, v_t \sim v_u)$, called the clustering coefficient, is positive and bounded away from zero. In particular, the underlying bipartite graph structure serves as a clustering mechanism.

Let us compare our model with the model of evolving network considered recently by Britton, Lindholm and Turova (2011) [6], (see also [5], [18] [19]). In their model vertices are prescribed weights, called social indices, and a vertex v_t with social index s_t creates new edges at a rate proportional to s_t . Clearly, both weight sequences $\{y_t\}_{t \geq 1}$ and $\{s_t\}_{t \geq 1}$ have the same purpose of modeling inhomogeneity of adjacency relations (hence both models possess a power law asymptotic degree distribution). But the model of Britton, Lindholm and Turova (2011)[6] does not have the clustering property. We remark, that the role of a bipartite structure in understanding/explaining clustering properties of some social networks has been discussed in Newman, Watts, and Strogatz (2002) [14]. Furthermore, empirically observed clustering properties of real affiliation networks have been reproduced with remarkable accuracy by related models of random intersection graphs, see [2], [3].

In the present paper we only consider the graph $G_{X,Y}$. We show the asymptotic distribution of the degree $d(v_t)$ of a vertex v_t as time $t \rightarrow +\infty$. We also obtain explicit asymptotic expressions for clustering coefficients $\alpha_{t|s,u}, \alpha_{s|t,u}, \alpha_{u|s,t}$, for $s, t, u \rightarrow +\infty$ such that $s < t < u$, and for the assortativity coefficient (Pearson's correlation coefficient between degrees of adjacent vertices)

$$r_{s,t} = \frac{\mathbf{E}_{st}d(v_s)d(v_t) - \mathbf{E}_{st}d(v_s)\mathbf{E}_{st}d(v_t)}{\sqrt{\mathbf{Var}_{st}d(v_s)\mathbf{Var}_{st}d(v_t)}}. \quad (2)$$

Here \mathbf{E}_{st} denotes the conditional expectation given the event $v_s \sim v_t$ and $\mathbf{Var}_{st}d(v_s) = \mathbf{E}_{st}d^2(v_s) - (\mathbf{E}_{st}d(v_s))^2$. We remark that (empirical) clustering and assortativity coefficients are commonly used characteristics of statistical dependence of adjacency relations of real networks.

Our results are stated in Section 2. Proofs are given in Section 3.

2 Results

Degree. We first present our results on the asymptotic degree distribution in $G_{X,Y}$. We obtain a compound probability distribution in the case where $\tau(t)$ grows linearly in t (clustering regime). For $\tau(t)$ growing faster than linearly in t , we obtain a mixed Poisson asymptotic degree distribution. We denote $a_k = \mathbf{E}X_1^k$, and $b_k = \mathbf{E}Y_1^k$.

Theorem 1. *Let $b > a > 0$. Let $\tau(t) = t$. Suppose that $\mathbf{E}X_1^2 < \infty$ and $\mathbf{E}Y_1 < \infty$. For $t \rightarrow +\infty$*

the random variable $d(v_t)$ converges in distribution to the random variable

$$d_* = \sum_{j=1}^{\Lambda_1} \varkappa_j, \quad (3)$$

where $\varkappa_1, \varkappa_2, \dots$ are independent and identically distributed random variables independent of the random variable Λ_1 . They are distributed as follows. For $r = 0, 1, 2, \dots$, we have

$$\mathbf{P}(\varkappa_1 = r) = \frac{r+1}{\mathbf{E}\Lambda_2} \mathbf{P}(\Lambda_2 = r+1) \quad \text{and} \quad \mathbf{P}(\Lambda_i = r) = \mathbf{E} e^{-\lambda_i} \frac{\lambda_i^r}{r!}, \quad i = 1, 2. \quad (4)$$

Here $\lambda_1 = 2(b^{1/2} - a^{1/2})a_1 Y_1$ and $\lambda_2 = 2(a^{-1/2} - b^{-1/2})b_1 X_1$.

The second moment condition $\mathbf{E}X_1^2 < \infty$ of Theorem 1 seems to be redundant.

Theorem 2. Let $b > a > 0$ and $\nu > 1$. Let $\tau(t) = t^\nu$, $t = 1, 2, \dots$. Suppose that $\mathbf{E}X_1^2 < \infty$ and $\mathbf{E}Y_1 < \infty$. For $t \rightarrow +\infty$ the random variable $d(v_t)$ converges in distribution to the random variable Λ_3 having the probability distribution

$$\mathbf{P}(\Lambda_3 = r) = \mathbf{E} e^{-\lambda_3} \frac{\lambda_3^r}{r!}, \quad r = 0, 1, 2, \dots \quad (5)$$

Here $\lambda_3 = \gamma a_2 b_1 Y_1$ and $\gamma = 4\nu(b^{1/2\nu} - a^{1/2\nu})(a^{-1/2\nu} - b^{-1/2\nu})$.

Remark 1. The result of Theorem 2 extends to a more general class of increasing nonnegative functions τ . In particular, assuming that

$$\lim_{t \rightarrow +\infty} \frac{t}{\tau(t)} = 0, \quad \sup_{t > 1} \frac{\tau^{-1}(2t)}{\tau^{-1}(t)} < \infty, \quad (6)$$

and that there exists finite limit

$$\gamma^* = \lim_{t \rightarrow +\infty} t^{-1/2} \sum_{a\tau(t) \leq i \leq b\tau(i)} i^{-1} \sum_{j: a\tau(j) \leq i \leq b\tau(j)} j^{-1/2},$$

we obtain the convergence in distribution of $d(v_t)$ to Λ_3 defined by (5) with $\lambda_3 = \gamma^* a_2 b_1 Y_1$. Here τ^{-1} denotes the inverse of τ (i.e., $\tau(\tau^{-1}(t)) = t$).

Remark 2. The function $\tau(t) = t \ln t$, which grows slower than any power t^ν , $\nu > 1$, satisfies conditions of Remark 1 with $\gamma^* = 4(a^{-1/2} - b^{-1/2})(b^{1/2} - a^{1/2})$. Furthermore, the functions $\tau_1(t) = e^{\ln^2 t}$ and $\tau_2(t) = e^t$, that grow faster than any power t^ν , satisfy conditions of Remark 1 with $\gamma^* = 0$.

Clustering. Our next result, Theorem 3, provides explicit asymptotic formulas for clustering coefficients. We note that for $s < t < u$ the conditional probabilities $\alpha_{s|tu}$, $\alpha_{t|su}$ and $\alpha_{u|st}$ are all different and, given $0 < a < b$, mainly depend on the ratios s/t , s/u and t/u . Denote $p_\Delta := p_\Delta(s, t, u) = \mathbf{P}(v_s \sim v_t, v_s \sim v_u, v_t \sim v_u)$ the probability that v_s, v_t, v_u make up a triangle.

Theorem 3. Let $b > a > 0$. Let $\tau(t) = t$. Suppose that $\mathbf{E}X_1^3 < \infty$ and $\mathbf{E}Y_1^2 < \infty$. Assume that $s, t, u \rightarrow +\infty$ so that $s < t < u$ and $\lceil au \rceil \leq \lfloor bs \rfloor$. We have

$$p_\Delta = \frac{a_3 b_1^3}{\sqrt{stu}} \left(\frac{2}{\sqrt{au}} - \frac{2}{\sqrt{bs}} \right) + o(t^{-2}), \quad (7)$$

$$\alpha_{t|su} = \frac{p_\Delta}{p_\Delta + a_2^2 b_1^2 b_2 t^{-1} (su)^{-1/2} \delta_{t|su}} + o(1), \quad (8)$$

$$\alpha_{s|tu} = \frac{p_\Delta}{p_\Delta + a_2^2 b_1^2 b_2 s^{-1} (tu)^{-1/2} \delta_{s|tu}} + o(1), \quad (9)$$

$$\alpha_{u|st} = \frac{p_\Delta}{p_\Delta + a_2^2 b_1^2 b_2 u^{-1} (st)^{-1/2} \delta_{u|st}} + o(1). \quad (10)$$

Here

$$\begin{aligned}\delta_{t|su} &= \ln(u/t) \ln(t/s) + \ln(u/t) \ln(bs/au) + \ln(t/s) \ln(bs/au) + \ln^2(bs/au), \\ \delta_{s|tu} &= \ln(u/t) \ln(bs/au) + \ln^2(bs/au), \\ \delta_{u|st} &= \ln(t/s) \ln(bs/au) + \ln^2(bs/au).\end{aligned}$$

We remark that the condition $\lceil au \rceil \leq \lfloor bs \rfloor$ of Theorem 3 excludes the trivial case where $p_\Delta \equiv 0$. Indeed, for $s < u$, the converse inequality $\lceil au \rceil > \lfloor bs \rfloor$ means that the lifetimes of v_s and v_u do not intersect and, therefore, we have $\mathbf{P}(v_s \sim v_u) \equiv 0$. In addition, the inequality $\lceil au \rceil \leq \lfloor bs \rfloor$ implies that positive numbers $\delta_{t|su}$, $\delta_{s|tu}$, $\delta_{u|st}$ are bounded from above by a constant (only depending on a and b).

Assortativity. Let us now consider the sequence of random variables $\{d(v_t)\}_{t \geq 1}$. We assume that $\tau(t) = t$. From Theorem 1 we know about the possible limiting distributions for $d(v_t)$. Moreover, from the fact that $G_{X,Y}$ is sparse we can conclude that, for any given k , the random variables $d(v_t), d(v_{t+1}), \dots, d(v_{t+k})$ are asymptotically independent as $t \rightarrow +\infty$. An interesting question is about the statistical dependence between $d(v_s)$ and $d(v_t)$ if we know, in addition, that vertices v_s and v_t are adjacent in $G_{X,Y}$. We assume that $s < t$ and let $s, t \rightarrow +\infty$ so that $bs - at \rightarrow +\infty$. Note that the latter condition ensures that the shared lifetime of v_s and v_t tends to infinity as $s, t \rightarrow +\infty$. In this case we obtain that conditional moments

$$\begin{aligned}\mathbf{E}_{st}d(v_s) &= \mathbf{E}_{st}d(v_t) + o(1) = \delta_1 + o(1), \\ \mathbf{E}_{st}d^2(v_s) &= \mathbf{E}_{st}d^2(v_t) + o(1) = \delta_2 + o(1), \\ \mathbf{E}_{st}d(v_s)d(v_t) &= \delta_2 - \Delta + o(1),\end{aligned}\tag{11}$$

are asymptotically constant. Here $\Delta = h_1^{-1}(2h_3 + 2h_5 + 4(h_6 - h_7))$ and

$$\delta_1 = 1 + h_1^{-1}(h_2 + 2h_3), \quad \delta_2 = 1 + h_1^{-1}(3h_2 + 6h_3 + h_4 + 6h_5 + 4h_6).$$

Furthermore, we denote

$$\begin{aligned}h_1 &= a_2 b_1^2, & h_2 &= a_3 b_1^3 \tilde{\gamma}, & h_3 &= a_2^2 b_1^2 b_2 \tilde{\gamma}(\sqrt{b} - \sqrt{a}), \\ h_4 &= a_4 b_1^4 \tilde{\gamma}^2, & h_5 &= a_2 a_3 b_1^3 b_2 \tilde{\gamma}^2(\sqrt{b} - \sqrt{a}), \\ h_6 &= a_2^3 b_1^3 b_3 \tilde{\gamma}^2(\sqrt{b} - \sqrt{a})^2, & h_7 &= a_2^3 b_1^2 b_2^2 \tilde{\gamma}^2(\sqrt{b} - \sqrt{a})^2,\end{aligned}\tag{12}$$

and $\tilde{\gamma} = 2(a^{-1/2} - b^{-1/2})$. A sketch of the derivation of relations (11) is given in Section 3 below. Finally, from (2) and (11) we obtain that the assortativity coefficient

$$r_{st} = 1 - \frac{\Delta}{\delta_2 - \delta_1^2} + o(1)\tag{13}$$

is asymptotically constant.

We note that each vertex of the graph $G_{X,Y}$ can be identified with the random subset of W , consisting of items selected by that vertex, and two vertices are adjacent in $G_{X,Y}$ whenever their subsets intersect. Graphs describing such adjacency relations between members of a *finite* family $\tilde{V} = \{\tilde{v}_1, \dots, \tilde{v}_n\}$ of random subsets of a given *finite* set $\tilde{W} = \{w_1, \dots, w_m\}$ are called random intersection graphs, see [13], [15] and [9]. Our graph $G_{X,Y}$ is, therefore, a random intersection graph evolving in time. One important application of random intersection graphs, defined by random subsets of fixed size, is the model of a secure wireless sensor network that

uses random predistribution of keys introduced in [8]. Another potential application of random intersection graphs is the statistical analysis and modeling of affiliation networks. For example, they are useful in explaining clustering properties of the actor network, see [2], [3]. Finally, we mention that asymptotic degree distribution and clustering properties of random intersection graphs have been studied in [1], [2], [7], [10], [11], [12], [16].

Concluding remarks. We have shown that the random graph $G_{X,Y}$ admits tunable asymptotic degree distribution (including the power law) and clustering and assortativity coefficients. An interesting problem were to study $G_{X,Y}$ and $\{G_{X,Y}(t)\}_{t \geq 1}$ in the case where deterministic cut-offs $a < b$ in (1) are replaced by random cut-offs $A_j \leq B_j$ (so that the lifetime $[A_j\tau(j), B_j\tau(j)]$ of an actor v_j were random). Furthermore, abrupt cut-offs can be replaced by some smooth cut-off functions.

3 Proofs

We first introduce some notation. Then we prove Theorems 1, 2, 3. The proof of Remark 1 goes along the lines of the proof of Theorem 2 and is omitted.

Throughout the proof limits are taken as $t \rightarrow +\infty$, if not stated otherwise. By c we denote positive numbers which may only depend on a, b and τ . We remark that c may attain different values in different places. We say that a sequence of random variables $\{\zeta_t\}_{t \geq 1}$ converges to zero in probability (denoted $\zeta_t = o_P(1)$) whenever $\limsup_t \mathbf{P}(|\zeta_t| > \varepsilon) = 0$ for each $\varepsilon > 0$. The sequence $\{\zeta_t\}_{t \geq 1}$ is called stochastically bounded (denoted $\zeta_t = O_P(1)$) whenever for each $\delta > 0$ there exists $N_\delta > 0$ such that $\limsup_t \mathbf{P}(|\zeta_t| > N_\delta) < \delta$.

Time intervals

$$T_t = \{i : a\tau(t) \leq i \leq b\tau(t)\}, \quad T_i^* = \{j : a\tau(j) \leq i \leq b\tau(j)\} \quad (14)$$

can be interpreted as lifetimes of the actor v_t and attribute w_i respectively. Here and below elements of V are called actors, elements of W are called attributes. The oldest and youngest actors that may establish a communication link with v_t are denoted v_{t_-} and v_{t_+} . Here

$$t_- = \min\{j : T_j \cap T_t \neq \emptyset\}, \quad t_+ = \max\{j : T_j \cap T_t \neq \emptyset\}.$$

The event "edge $\{w_i, v_j\}$ is present in $H_{X,Y}$ " is denoted $w_i \rightarrow v_j$. Introduce random variables

$$\begin{aligned} \mathbb{I}_{ij} &= \mathbb{I}_{\{w_i \rightarrow v_j\}}, & \mathbb{I}_i &= \mathbb{I}_{it}, & u_i &= \sum_{j \in T_i^* \setminus \{t\}} \mathbb{I}_{ij}, & L &= L_t = \sum_{i \in T_t} u_i \mathbb{I}_i, \\ b_k(I) &= \sum_{j \in I} Y_j^k j^{-k/2}, & a_k(I) &= \sum_{i \in I} X_i^k i^{-k/2}, & I &\subset \mathbb{N}, \end{aligned} \quad (15)$$

$$\lambda_{ij} = X_i Y_j / \sqrt{ij}, \quad Q_{XY}(t) = \sum_{i \in T_t} \lambda_{it} \sum_{j \in T_i^* \setminus \{t\}} \lambda_{ij} \min\{1, \lambda_{ij}\}. \quad (16)$$

We remark, that u_i counts all neighbours of w_i in $H_{X,Y}$ belonging to the set $V \setminus \{v_t\}$, and L_t counts all paths of length 2 in $H_{X,Y}$ starting from v_t . Introduce events

$$\mathcal{A}_t = \{\lambda_{it} \leq 1, i \in T_t\}, \quad \mathcal{B}_t(\varepsilon) = \{Y_j \leq \varepsilon^2 j, j \in [t_-, t_+] \setminus \{t\}\}, \quad \varepsilon > 0.$$

By $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{E}}$ we denote the conditional probability and expectation given X, Y . The conditional probability and expectation given Y is denoted \mathbf{P}_Y and \mathbf{E}_Y . By \mathbf{P}_t and \mathbf{E}_t we denote the conditional probability and expectation given Y_t . By $d_{TV}(\zeta, \xi)$ we denote the total variation

distance between the probability distributions of random variables ζ and ξ . In the case where ζ, ξ and X, Y are defined on the same probability space, we denote by $\tilde{d}_{TV}(\zeta, \xi)$ the total variation distance between the conditional distributions of ζ and ξ given X, Y .

In the proof we use the following simple fact. For a uniformly bounded sequence of random variables $\{\zeta_t\}_{t \geq 1}$ (i.e., \exists nonrandom $h > 0$ such that $\forall t \ 0 < \zeta_t < h$ almost surely) we have

$$\zeta_t = o_P(1) \Rightarrow \mathbf{E}\zeta_t = o(1). \quad (17)$$

In particular, given a sequence of bivariate random vectors $\{(\phi_t, \psi_t)\}_{t \geq 1}$, defined on the same probability space as X, Y , we have

$$\tilde{d}_{TV}(\phi_t, \psi_t) = o_P(1) \Rightarrow d_{TV}(\phi_t, \psi_t) = o(1). \quad (18)$$

3.1 Proof of Theorem 1

Before the proof we collect auxiliary results. For $\tau(t) := t$ and T_t, T_i^* defined in (14), we have

$$\begin{aligned} \sum_{i \in T_t} i^{-1/2} &= t^{1/2} \gamma_1 + rt^{-1/2}, & \sum_{j \in T_i^*} j^{-1/2} &= i^{1/2} \gamma_2 + r'i^{-1/2}, \\ \gamma_1 &:= 2(b^{1/2} - a^{1/2}), & \gamma_2 &:= 2(a^{-1/2} - b^{-1/2}), \end{aligned} \quad (19)$$

where $|r|, |r'| \leq c$.

Lemma 1. *Let $t \rightarrow +\infty$. Assume that $\mathbf{E}X_1^2 < \infty$ and $\mathbf{E}Y_1 < \infty$. We have*

$$\forall \varepsilon > 0 \quad \mathbf{P}(\mathcal{B}_t(\varepsilon)) = 1 - o(1), \quad (20)$$

$$t^{-1}b_2([t_-, t_+] \setminus \{t\}) = o_P(1), \quad (21)$$

$$\mathbf{P}(d(v_t) \neq L_t) = o(1), \quad (22)$$

$$\mathbf{P}(\mathcal{A}_t) = 1 - o(1), \quad (23)$$

$$Q_{XY}(t) = o_P(1), \quad \mathbf{E}Q_{XY}(t) = o(1). \quad (24)$$

For any integers $t > a^{-1}(b + b^{-1})$ and $i \in T_t$, and any $0 < \varepsilon < 1$ we have

$$|\mathbf{E}a_1(T_t) - a_1 \gamma_1 t^{1/2}| \leq ca_1 t^{-1/2}, \quad |\mathbf{E}b_1(T_i^* \setminus \{t\}) - b_1 \gamma_2 i^{1/2}| \leq cb_1 i^{-1/2}, \quad (25)$$

$$\mathbf{E}|b_1(T_i^* \setminus \{t\}) - b_1 \gamma_2 i^{1/2}| \mathbb{I}_{\mathcal{B}_t(\varepsilon)} \leq ci^{1/2}(\varepsilon b_1^{1/2} + \mathbf{E}Y_1 \mathbb{I}_{\{Y_1 > \varepsilon^2 t_-\}}) + cb_1 i^{-1/2}, \quad (26)$$

$$\mathbf{E}|a_1(T_t) - a_1 \gamma_1 t^{1/2}| \leq ca_2^{1/2}. \quad (27)$$

Proof of Lemma 1. Proof of (20). We estimate the probability of the complement event $\bar{\mathcal{B}}_t(\varepsilon)$ using the union bound and Markov's inequality

$$\mathbf{P}(\bar{\mathcal{B}}_t(\varepsilon)) \leq \sum_{t_- \leq j \leq t_+} \mathbf{P}(Y_i > \varepsilon^2 j) = t_+ \mathbf{P}(Y_1 > \varepsilon^2 t_-) \leq \varepsilon^{-2} (t_+/t_-) \mathbf{E}Y_1 \mathbb{I}_{\{Y_1 > \varepsilon^2 t_-\}} = o(1).$$

Here we estimate $t_+/t_- \leq c$ and invoke the bound $\mathbf{E}Y_1 \mathbb{I}_{\{Y_1 > s\}} = o(1)$, for $s \rightarrow +\infty$.

Proof of (21). Denote $\hat{b}_2(t) = t^{-2} \sum_{1 \leq j \leq t} Y_j^2$. We note that $\mathbf{E}Y_1 < \infty$ implies $\hat{b}_2(t) = o_P(1)$. The latter bound in combination with the simple inequality $t_+/t_- \leq c$ implies (21).

Proof of (23). Let $\bar{\mathcal{A}}_t$ denote the complement event to \mathcal{A}_t . We have, by the union bound and Markov's inequality,

$$\mathbf{P}_t(\bar{\mathcal{A}}_t) \leq \sum_{i \in T_t} \mathbf{P}_t(\lambda_{it} \geq 1) \leq \sum_{i \in T_t} (it)^{-1} Y_t^2 a_2 \leq ca_2 t^{-1} Y_t^2.$$

We obtain the bound $\mathbf{P}_t(\overline{\mathcal{A}}_1) = o(1)$, which implies (23), see (17).

Proof of (22). In view of (17) it suffices to show that $\mathbf{P}_X(d(v_t) \neq L_t) = o_P(1)$. We note that $d(v_t) \neq L_t$ if and only if $S \geq 1$, where $S = \sum' \mathbb{I}_{i_1} \mathbb{I}_{i_2} \mathbb{I}_{i_1 j} \mathbb{I}_{i_2 j}$. Here we denote $\sum' = \sum_{\{i_1, i_2\} \subset T_t} \sum_{j \in T_{i_1}^* \cap T_{i_2}^*, j \neq t}$. Observing that

$$\mathbf{E}_X \mathbb{I}_{i_1} \mathbb{I}_{i_2} \mathbb{I}_{i_1 j} \mathbb{I}_{i_2 j} = \mathbf{E}_X p_{i_1 t} p_{i_2 t} p_{i_1 j} p_{i_2 j} \leq a_2^2 Y_t^2 / (i_1 i_2 j)^{-1}$$

we obtain, by Markov's inequality,

$$\mathbf{P}_X(d(v_t) \neq L_t) = \mathbf{P}_X(S \geq 1) \leq \mathbf{E}_X S \leq a_2^2 Y_t^2 t^{-1} \sum' Y_j^2 (i_1 i_2 j)^{-1}. \quad (28)$$

The simple bound $\sum_{\{i_1, i_2\} \subset T_t} \frac{1}{i_1 i_2} \leq c$ implies $\sum' Y_j^2 (i_1 i_2 j)^{-1} \leq c b_2([t_-, t_+] \setminus \{t\})$. Now, by (21) the right-hand side of (28) tends to zero in probability.

Proof of (24). Denote $\hat{X}_i = \max\{X_i, 1\}$, $\hat{Y}_j = \max\{Y_j, 1\}$, and let $\hat{Q}_{XY}(t)$ denote the sum (16), where λ_{ij} is replaced by $\hat{\lambda}_{ij} = \hat{X}_i \hat{Y}_j / \sqrt{i j}$. We observe that $\mathbf{E} X_1^2 < \infty$ and $\mathbf{E} Y_1 < \infty$ imply

$$a_\varphi := \mathbf{E} \hat{X}_1^2 \varphi(\hat{X}_1) < \infty, \quad b_\varphi := \mathbf{E} \hat{Y}_1 \varphi(\hat{Y}_1) < \infty,$$

for some positive increasing function $\varphi : [1, +\infty) \rightarrow [0, +\infty)$ satisfying $\varphi(u) \rightarrow +\infty$ as $u \rightarrow +\infty$ (clearly, $\varphi(\cdot)$ depends on the distributions of X_1 and Y_1). In addition, we can choose φ satisfying $\varphi(u) \leq u$ and $\varphi(su) \leq \varphi(s)\varphi(u)$, for $s, u \geq 1$. From these inequalities one derives the inequality $\min\{1, \hat{\lambda}_{ij}\} \leq \varphi(\hat{X}_i) \varphi(\hat{Y}_j) / \varphi(\sqrt{i j})$. The latter inequality implies

$$\hat{Q}_{XY}(t) \leq \hat{Y}_t Q_{XY}^*(t), \quad Q_{XY}^*(t) := \sum_{i \in T_t} \frac{\hat{X}_i^2 \varphi(\hat{X}_i)}{\sqrt{ti}} \sum_{j \in T_i^* \setminus \{t\}} \frac{\hat{Y}_j \varphi(\hat{Y}_j)}{\sqrt{ij}} \frac{1}{\varphi(\sqrt{ij})}$$

Furthermore, for $i \in T_t$ and $j \in T_i^*$ we have $ij \geq \lfloor at \rfloor t_- =: t_*^2$, and $t_* \rightarrow +\infty$ as $t \rightarrow +\infty$. Hence

$$\mathbf{E} Q_{XY}^*(t) \leq \frac{1}{\varphi(t_*)} a_\varphi b_\varphi \sum_{i \in T_t} \frac{1}{\sqrt{ti}} \sum_{j \in T_i^*} \frac{1}{\sqrt{ij}} = O\left(\frac{1}{\varphi(t_*)}\right) = o(1).$$

This bound together with the inequalities $Q_{XY}(t) \leq \hat{Q}_{XY}(t) \leq \hat{Y}_t Q_{XY}^*(t)$ shows (24).

Proof of (25). These inequalities follow from (19).

Proof of (26). Denote $\Delta = |\tilde{b} - b_1 \gamma_2 i^{1/2}|$, where \tilde{b} denotes the sum $b_1(T_i^* \setminus \{t\})$, but with Y_j replaced by $\tilde{Y}_j = Y_j \mathbb{I}_{\{Y_j \leq \varepsilon^2 j\}}$, $j \in T_i^* \setminus \{t\}$. We have

$$|b_1(T_i^* \setminus \{t\}) - b_1 \gamma_2 i^{1/2}| \mathbb{I}_{\mathcal{B}_t(\varepsilon)} = \Delta \mathbb{I}_{\mathcal{B}_t(\varepsilon)} \leq \Delta \leq \Delta_1 + \Delta_2 + \Delta_3, \quad (29)$$

where we denote $\Delta_1 = |\tilde{b} - \mathbf{E} \tilde{b}|$, $\Delta_2 = |\mathbf{E} \tilde{b} - \mathbf{E} b_1(T_i^* \setminus \{t\})|$, $\Delta_3 = |\mathbf{E} b_1(T_i^* \setminus \{t\}) - b_1 \gamma_2 i^{1/2}|$. Next, we evaluate $\mathbf{E} \Delta_1$ and Δ_2 :

$$(\mathbf{E} \Delta_1)^2 \leq \mathbf{E} (\Delta_1^2) \leq \sum_{j \in T_i^* \setminus \{t\}} j^{-1} \mathbf{E} \tilde{Y}_j^2 \leq \varepsilon^2 b_1 |T_i^*|, \quad (30)$$

$$\Delta_2 \leq \sum_{j \in T_i^* \setminus \{t\}} j^{-1/2} \mathbf{E} \tilde{Y}_j \mathbb{I}_{\{Y_j > \varepsilon^2 j\}} \leq \mathbf{E} Y_1 \mathbb{I}_{\{Y_1 > \varepsilon^2 t_-\}} \sum_{j \in T_i^* \setminus \{t\}} j^{-1/2}. \quad (31)$$

In (30) we first apply Cauchy-Schwartz, then use the linearity of variance of an iid sum, and finally apply the inequality $\mathbf{Var} \tilde{Y}_j \leq \mathbf{E} \tilde{Y}_j^2 \leq j^{-1} \varepsilon^2 \mathbf{E} Y_j$. Invoking (25), (30), (31) in (29) and using (19) and $|T_i^*| \leq ci$ we obtain (26).

Proof of (27). We write $\mathbf{E}|a_1(T_t) - a_1\gamma_1 t^{1/2}| \leq \mathbf{E}\tilde{\Delta}_1 + \tilde{\Delta}_2$, where $\tilde{\Delta}_1 := |a_1(T_t) - \mathbf{E}a_1(T_t)|$ and $\tilde{\Delta}_2 = |\mathbf{E}a_1(T_t) - a_1\gamma_1 t^{1/2}|$, and invoke the inequalities

$$(\mathbf{E}\tilde{\Delta}_1)^2 \leq \mathbf{E}\tilde{\Delta}_1^2 = \sum_{i \in T_t} j^{-1}(a_2 - a_1^2) \leq ca_2$$

and $\tilde{\Delta}_2 \leq ca_1 \leq ca_2^{1/2}$, see (25). \square

Inequality (32) below is referred to as LeCam's inequality, see e.g., [17].

Lemma 2. *Let $S = \mathbb{I}_1 + \mathbb{I}_2 + \dots + \mathbb{I}_n$ be the sum of independent random indicators with probabilities $\mathbf{P}(\mathbb{I}_i = 1) = p_i$. Let Λ be Poisson random variable with mean $p_1 + \dots + p_n$. The total variation distance between the distributions P_S of P_Λ of S and Λ*

$$d_{TV}(S, \Lambda) := \sup_{A \subset \{0, 1, 2, \dots\}} |\mathbf{P}(S \in A) - \mathbf{P}(\Lambda \in A)| \leq \sum_i p_i^2. \quad (32)$$

Proof of Theorem 1. Before the proof we introduce some notation. Given X, Y , we generate independent Poisson random variables

$$\eta_i, \quad \xi_{1i}, \quad \xi_{3i}, \quad \xi_{4i}, \quad \Delta_{ri}, \quad i \in T_t, \quad r = 1, 2, 3,$$

with conditional mean values

$$\begin{aligned} \tilde{\mathbf{E}}\eta_i &= \lambda_{it}, & \tilde{\mathbf{E}}\xi_{1i} &= \sum_{j \in T_i^* \setminus \{t\}} p_{ij}, & \tilde{\mathbf{E}}\xi_{3i} &= X_i b_1 \gamma_2, & \tilde{\mathbf{E}}\xi_{4i} &= X_i \bar{b} i^{-1/2}, \\ \tilde{\mathbf{E}}\Delta_{1i} &= \sum_{j \in T_i^* \setminus \{t\}} (\lambda_{ij} - p_{ij}), & \tilde{\mathbf{E}}\Delta_{2i} &= X_i \delta_{2i} i^{-1/2}, & \tilde{\mathbf{E}}\Delta_{3i} &= X_i \delta_{3i} i^{-1/2}. \end{aligned}$$

Here

$$\delta_{2i} = b_1(T_i^* \setminus \{t\}) - \bar{b}, \quad \delta_{3i} = b_1 \gamma_2 i^{1/2} - \bar{b}, \quad \bar{b} = \min\{b_1(T_i^* \setminus \{t\}), b_1 \gamma_2 i^{1/2}\}.$$

Finally, we define $\xi_{2i} = \xi_{1i} + \Delta_{1i}$, $i \in T_t$ and introduce random variables

$$L_{0t} = \sum_{i \in T_t} \eta_i u_i, \quad L_{rt} = \sum_{i \in T_t} \eta_i \xi_{ri}, \quad r = 1, 2, 3. \quad (33)$$

We assume, in addition, that given X, Y the families of random variables $\{\mathbb{I}_i, i \in T_t\}$ and $\{\xi_{ri}, i \in T_t, r = 1, 2, 3, 4\}$ are conditionally independent, and that $\{\eta_i, i \in T_t\}$ is conditionally independent of the set of edges of $H_{X,Y}$ that are not incident to v_t .

We are ready to start the proof. In view of (22) the random variables $d(v_t)$ and L_t have the same asymptotic distribution (if any). We shall prove that L_t converges in distribution to d_* . In the proof we approximate L_t by the random variable L_{3t} , see (34) and (35) below. Afterwards we show that L_{3t} converges in distribution to d_* .

In order to show that L_t and L_{3t} have the same asymptotic distribution (if any) we prove the bounds

$$d_{TV}(L_t, L_{0t}) = o(1), \quad d_{TV}(L_{0t}, L_{1t}) = o(1), \quad (34)$$

$$\mathbf{E}|L_{1t} - L_{2t}| = o(1), \quad \tilde{L}_{2t} - \tilde{L}_{3t} = o_P(1). \quad (35)$$

Here \tilde{L}_{2t} and \tilde{L}_{3t} are marginals of the random vector $(\tilde{L}_{2t}, \tilde{L}_{3t})$ constructed in (39) below which has the property that \tilde{L}_{2t} has the same distribution as L_{2t} and \tilde{L}_{3t} has the same distribution as L_{3t} .

Let us prove the first bound of (34). We shall show below that

$$\tilde{d}_{TV}(L_t, L_{0t}) \mathbb{I}_{\mathcal{A}_t} \leq t^{-1} Y_t^2 a_2(T_t). \quad (36)$$

From the inequality $\mathbf{E} a_2(T_t) = \sum_{i \in T_t} i^{-1} a_2 \leq c a_2$ we conclude that $a_2(T_t)$ is stochastically bounded. Hence $t^{-1} Y_t^2 a_2(T_t) = o_P(1)$. This bound and (36) combined with (23) imply

$$\tilde{d}_{TV}(L_t, L_{0t}) \leq \tilde{d}_{TV}(L_t, L_{0t}) \mathbb{I}_{\mathcal{A}_t} + \mathbb{I}_{\overline{\mathcal{A}}_t} = o_P(1).$$

Now the first bound of (34) follows from (18). It remains to prove (36). We denote $L'_k = \sum_{i=\lfloor at \rfloor}^k \mathbb{I}_i u_i + \sum_{i=k+1}^{\lfloor bt \rfloor} \eta_i u_i$ and write, by the triangle inequality,

$$\tilde{d}_{TV}(L_t, L_{0t}) \leq \sum_{k \in T_t} \tilde{d}_{TV}(L'_{k-1}, L'_k).$$

Then we estimate $\tilde{d}_{TV}(L'_{k-1}, L'_k) \leq \tilde{d}_{TV}(\eta_k, \mathbb{I}_k) \leq (kt)^{-1} Y_t^2 X_k^2$. Here the first inequality follows from the properties of the total variation distance. The second inequality follows from Lemma 2 and the fact that on the event \mathcal{A}_t we have $p_{kt} = \lambda_{kt}$.

Let us prove the second bound of (34). In view of (17) it suffices to show that $\tilde{d}_{TV}(L_{0t}, L_{1t}) = o_P(1)$. For this purpose we write, by the triangle inequality,

$$\tilde{d}_{TV}(L_{0t}, L_{1t}) \leq \sum_{k \in T_t} \tilde{d}_{TV}(L^*_{k-1}, L^*_k), \quad (37)$$

where $L^*_k := \sum_{i=\lfloor at \rfloor}^k \eta_i u_i + \sum_{i=k+1}^{\lfloor bt \rfloor} \eta_i \xi_{1i}$, and estimate

$$\tilde{d}_{TV}(L^*_{k-1}, L^*_k) \leq \tilde{d}_{TV}(\eta_k u_k, \eta_k \xi_{1k}) \leq \tilde{\mathbf{P}}(\eta_k \neq 0) \tilde{d}_{TV}(u_k, \xi_{1k}). \quad (38)$$

Now, invoking the inequalities

$$\tilde{\mathbf{P}}(\eta_k \neq 0) = 1 - e^{-\lambda_{kt}} \leq \lambda_{kt} \quad \text{and} \quad \tilde{d}_{TV}(u_k, \xi_{1k}) \leq \sum_{j \in T_k^* \setminus \{t\}} p_{kj}^2,$$

see (32), we obtain from (37), (38) and (24) that

$$\tilde{d}_{TV}(L_0, L_1) \leq Q_{XY}(t) = o_P(1).$$

Let us prove the first bound of (35). We observe that

$$|L_{2t} - L_{1t}| = L_{2t} - L_{1t} = \sum_{i \in T_t} \eta_i \Delta_{1i}$$

and

$$\tilde{\mathbf{E}} \sum_{i \in T_t} \eta_i \Delta_{1i} = \sum_{i \in T_t} \lambda_{it} \sum_{j \in T_i^* \setminus \{t\}} (\lambda_{ij} - 1) \mathbb{I}_{\{\lambda_{ij} > 1\}} \leq Q_{XY}(t).$$

We obtain $\mathbf{E}|L_{2t} - L_{1t}| \leq \mathbf{E}Q_{XY}(t) = o(1)$, see (24).

Let us prove the second bound of (35). We note that the random vector

$$(\tilde{L}_{2t}, \tilde{L}_{3t}), \quad \tilde{L}_{2t} = \sum_{i \in T_t} \eta_i (\xi_{4i} + \Delta_{2i}), \quad \tilde{L}_{3t} = \sum_{i \in T_t} \eta_i (\xi_{4i} + \Delta_{3i}) \quad (39)$$

has the marginal distributions of (L_{2t}, L_{3t}) . In addition, since $\Delta_{2i}, \Delta_{3i} \geq 0$ and at most one of them is non-zero, we have $|\Delta_{2i} - \Delta_{3i}| = \Delta_{2i} + \Delta_{3i}$. Therefore, we can write

$$\tilde{\Delta} := |\tilde{L}_{2t} - \tilde{L}_{3t}| \leq \sum_{i \in T_t} |\eta_i| |\Delta_{2i} - \Delta_{3i}| = \sum_{i \in T_t} \eta_i (\Delta_{2i} + \Delta_{3i}). \quad (40)$$

We remark that given X, Y the random variable $\Delta_{2i} + \Delta_{3i}$ has Poisson distribution with (conditional) mean value

$$\tilde{\mathbf{E}}(\Delta_{2i} + \Delta_{3i}) = X_i i^{-1/2} \delta_i, \quad \delta_i := |b_1(T_i^* \setminus \{t\}) - b_1 \gamma_2 i^{1/2}|.$$

Therefore, (40) implies $\tilde{\mathbf{E}}\tilde{\Delta} \leq t^{-1/2} Y_t \sum_{i \in T_t} X_i^2 i^{-1} \delta_i$. Next, for $0 < \varepsilon < 1$, we write

$$\mathbf{E}\mathbb{I}_{\mathcal{B}_t(\varepsilon)}\tilde{\Delta} \leq \mathbf{E}\mathbb{I}_{\mathcal{B}_t(\varepsilon)} t^{-1/2} Y_t \sum_{i \in T_t} X_i^2 i^{-1} \delta_i = b_1 a_2 t^{-1/2} \sum_{i \in T_t} i^{-1} \mathbf{E}\delta_i \mathbb{I}_{\mathcal{B}_t(\varepsilon)}.$$

Invoking upper bound (26) for $\mathbf{E}\delta_i \mathbb{I}_{\mathcal{B}_t(\varepsilon)}$ we obtain $\mathbf{E}\mathbb{I}_{\mathcal{B}_t(\varepsilon)}\tilde{\Delta} \leq cb_1^{3/2} a_2 \varepsilon + o(1)$. Finally, this bound combined with Markov's inequality and (20) yields

$$\mathbf{P}(\tilde{\Delta} \geq 1) = \mathbf{P}(\{\tilde{\Delta} \geq 1\} \cap \mathcal{B}_t(\varepsilon)) + o(1) \leq \mathbf{E}\mathbb{I}_{\mathcal{B}_t(\varepsilon)}\tilde{\Delta} + o(1) \leq cb_1^{3/2} a_2 \varepsilon + o(1).$$

We conclude that $\mathbf{P}(\tilde{\Delta} \neq 0) = \mathbf{P}(\tilde{\Delta} \geq 1) = o(1)$.

Next we prove that L_{3t} converges in distribution to d_* defined by (3). Let Y_* be a random copy of Y_1 , which is independent of X, Y . Given X, Y, Y_* , we generate independent Poisson random variables $\eta_k^*, k \in T_t$ with (conditional) mean values $\mathbf{E}(\eta_k^* | X, Y, Y_*) = \lambda_{k*}$, where $\lambda_{k*} = X_k Y_*(kt)^{-1/2}$. We assume that, given X, Y, Y_* , the family of random variables $\{\eta_k^*, k \in T_t\}$ is conditionally independent of $\{\xi_{3k}, k \in T_t\}$. Define $L_t^* = \sum_{k \in T_t} \eta_k^* \xi_{3k}$. We note that L_t^* is defined in the same way as L_{3t} above, but with Y_t replaced by Y_* . Let d_* be defined in the same way as d_* , but with λ_1 replaced by $\lambda_* = Y_* a_1 \gamma_1$. Since L_{3t} has the same distribution as L_t^* , and d_* has the same distribution as d_* , it suffices to show that L_t^* converges in distribution to d_* . For this purpose we show the convergence of Fourier-Stieltjes transforms $\mathbf{E}e^{izL_t^*} \rightarrow \mathbf{E}e^{izd_*}$, for each $z \in (-\infty, +\infty)$. Denote $\Delta^*(z) = e^{izL_t^*} - e^{izd_*}$. We shall show below that, for any real z and any realized value Y_* there exists a positive constant $c^* = c^*(z, Y_*)$ such that for every $0 < \varepsilon < 0.5$ we have

$$\limsup_t |\mathbf{E}(\Delta^*(z) | Y_*)| < c^* \varepsilon. \quad (41)$$

Clearly, (41) implies $\mathbf{E}(\Delta^*(z) | Y_*) = o(1)$. This fact together with the simple inequality $|\Delta^*(z)| \leq 2$ yields $\mathbf{E}\Delta^*(z) = o(1)$, by Lebesgue's dominated convergence theorem. Finally, the identity $\mathbf{E}\Delta^*(z) = \mathbf{E}e^{izL_t^*} - \mathbf{E}e^{izd_*}$ implies $\mathbf{E}e^{izL_t^*} \rightarrow \mathbf{E}e^{izd_*}$.

We fix $0 < \varepsilon < 0.5$ and prove (41). Before the proof we introduce some notation. Denote

$$f_\varkappa(z) = \mathbf{E}e^{iz\varkappa_1}, \quad \bar{f}_\varkappa(z) = \sum_{r \geq 0} e^{izr} \bar{p}_r, \quad \bar{p}_r = \bar{\lambda}^{-1} \sum_{k \in T_t} \lambda_{k*} \mathbb{I}_{\{\xi_{3k}=r\}}, \quad \bar{\lambda} = \sum_{k \in T_t} \lambda_{k*},$$

$$\delta = (\bar{f}_\varkappa(z) - 1)\bar{\lambda} - (f_\varkappa(z) - 1)\lambda_*, \quad f(z) = \mathbf{E}_* e^{izd_*}, \quad \bar{f}(z) = \bar{\mathbf{E}} e^{izL_t^*}.$$

Here $\bar{\mathbf{E}}$ denotes the conditional expectation given X, Y, Y_\star and $\{\xi_{3k}, k \in T_t\}$. By \mathbf{E}_\star we denote the conditional expectation given Y_\star .

Introduce the event $\mathcal{D} = \{|a_1(T_t) - \gamma_1 a_1 t^{1/2}| < \varepsilon t^{1/2} \min\{1, \gamma_1 a_1\}\}$ and let $\bar{\mathcal{D}}$ denote the complement event. Furthermore, select the number $T > 1/\varepsilon$ such that $\mathbf{P}(\varkappa_1 \geq T) < \varepsilon$. By $c_1^\star, c_2^\star, \dots$ we denote positive numbers which do not depend on t .

We observe that, given Y_\star , the conditional distribution of d_\star is the compound Poisson distribution with the characteristic function $f(z) = e^{\lambda_\star(f_\varkappa(z)-1)}$. Similarly, given X, Y, Y_\star and $\{\xi_{3k}, k \in T_t\}$, the conditional distribution of L_t^\star is the compound Poisson distribution with the characteristic function $\bar{f}(z) = e^{\bar{\lambda}(\bar{f}_\varkappa(z)-1)}$. In the proof of (41) we exploit the convergence $\bar{\lambda} \rightarrow \lambda_\star$ and $\bar{f}_\varkappa(z) \rightarrow f_\varkappa(z)$.

Let us prove (41). We write

$$\mathbf{E}_\star \Delta^\star(z) = I_1 + I_2, \quad I_1 = \mathbf{E}_\star \Delta^\star(z) \mathbb{I}_{\mathcal{D}}, \quad I_2 = \mathbf{E}_\star \Delta^\star(z) \mathbb{I}_{\bar{\mathcal{D}}}.$$

Here $|I_2| \leq 2\mathbf{P}_\star(\bar{\mathcal{D}}) = 2\mathbf{P}(\bar{\mathcal{D}}) = o(1)$. Indeed, the bound $\mathbf{P}(\bar{\mathcal{D}}) = o(1)$ follows from (27), by Markov's inequality. Next we estimate I_1 . Combining the identity $\mathbf{E}_\star \Delta^\star(z) = \mathbf{E}_\star f(z)(e^\delta - 1)$ with the inequalities $|f(z)| \leq 1$ and $|e^s - 1| \leq |s|e^{|s|}$, we obtain

$$|I_1| \leq \mathbf{E}_\star |\delta| e^{|\delta|} \mathbb{I}_{\mathcal{D}} \leq c_1^\star \mathbf{E}_\star |\delta| \mathbb{I}_{\mathcal{D}}. \quad (42)$$

Here we estimated $e^{|\delta|} \leq e^{6\lambda_\star} =: c_1^\star$ using the inequalities

$$|\delta| \leq 2\bar{\lambda} + 2\lambda_\star, \quad \bar{\lambda} = Y_\star t^{-1/2} a_1(T_t) \leq 2\lambda_\star.$$

We remark that the last inequality holds provided that event \mathcal{D} occurs.

Finally, we show that $\mathbf{E}_\star |\delta| \mathbb{I}_{\mathcal{D}} \leq (c_2^\star + c_3^\star \lambda_\star + c_4^\star \lambda_\star) \varepsilon + o(1)$. To this aim we write

$$\delta = (\bar{f}_\varkappa(z) - 1)(\bar{\lambda} - \lambda_\star) + (\bar{f}_\varkappa(z) - f_\varkappa(z))\lambda_\star$$

and estimate $|\delta| \leq 2|\bar{\lambda} - \lambda_\star| + \lambda_\star |\bar{f}_\varkappa(z) - f_\varkappa(z)|$. The inequality, which holds on the event \mathcal{D} , $|\bar{\lambda} - \lambda_\star| \leq Y_\star \varepsilon$ implies $\mathbf{E}_\star |\bar{\lambda} - \lambda_\star| \mathbb{I}_{\mathcal{D}} \leq c_2^\star \varepsilon$ with $c_2^\star := Y_\star$. Next we show that

$$\mathbf{E}_\star |\bar{f}_\varkappa(z) - f_\varkappa(z)| \mathbb{I}_{\mathcal{D}} \leq (c_3^\star + c_4^\star) \varepsilon + o(1).$$

We first split

$$\bar{f}_\varkappa(z) - f_\varkappa(z) = \sum_{r \geq 0} e^{izr} (\bar{p}_r - p_r) = R_1 - R_2 + R_3,$$

and then estimate separately the terms

$$R_1 = \sum_{r \geq T} e^{izr} \bar{p}_r, \quad R_2 = \sum_{r \geq T} e^{izr} p_r, \quad R_3 = \sum_{0 \leq r < T} e^{izr} (\bar{p}_r - p_r).$$

Here we denote $p_r = \mathbf{P}(\varkappa_1 = r)$. The upper bound for R_2 follows by the choice of T

$$|R_2| \leq \sum_{r \geq T} p_r = \mathbf{P}(\varkappa_1 \geq T) < \varepsilon.$$

Next, combining the identities $\bar{p}_r = (a_1(T_t))^{-1} \sum_{k \in T_t} k^{-1/2} X_k \mathbb{I}_{\{\xi_{3k}=r\}}$ and

$$\sum_{r \geq T} \sum_{k \in T_t} k^{-1/2} X_k \mathbb{I}_{\{\xi_{3k}=r\}} = \sum_{k \in T_t} k^{-1/2} X_k \mathbb{I}_{\{\xi_{3k} \geq T\}}$$

with the inequality $a_1(T_t) \geq t^{1/2}a_1\gamma_1/2$, which holds on the event \mathcal{D} , we obtain

$$|R_1|\mathbb{I}_{\mathcal{D}} \leq \sum_{r \geq T} \bar{p}_r \leq \frac{2}{a_1\gamma_1 t^{1/2}} \sum_{k \in T_t} \frac{X_k}{k^{1/2}} \mathbb{I}_{\{\xi_{3k} \geq T\}} \leq \frac{2}{a_1\gamma_1 t^{1/2}} \sum_{k \in T_t} \frac{X_k \xi_{3k}}{Tk^{1/2}}.$$

Now, the identity $\mathbf{E}_* X_k \xi_{3k} = a_2 b_1 \gamma_2$ implies $\mathbf{E}_* |R_1| \mathbb{I}_{\mathcal{D}} \leq c_4^* T^{-1} \leq c_4^* \varepsilon$.

Now we estimate R_3 . We denote $p'_r = a_1(T_t)(a_1\gamma_1 t^{1/2})^{-1} \bar{p}_r$ and observe that the inequality $|a_1(T_t)(a_1\gamma_1 t^{1/2})^{-1} - 1| \leq \varepsilon$, which holds on the event \mathcal{D} , implies

$$\left| \sum_{0 \leq r \leq T} e^{itr} (\bar{p}_r - p'_r) \right| \mathbb{I}_{\mathcal{D}} \leq \varepsilon \sum_{0 \leq r \leq T} \bar{p}_r \leq \varepsilon.$$

In the last inequality we use the fact that the probabilities $\{\bar{p}_r\}_{r \geq 0}$ sum up to 1. It follows now that

$$|R_3| \mathbb{I}_{\mathcal{D}} \leq \varepsilon + \sum_{0 \leq r \leq T} |p'_r - p_r|. \quad (43)$$

Next we estimate

$$\mathbf{E}_* |p'_r - p_r| \leq \mathbf{E}_* |p'_r - \mathbf{E}_* p'_r| + |\mathbf{E}_* p'_r - p_r| \quad (44)$$

where, by the Cauchy-Schwartz and the linearity of the variance of an iid sum, we have

$$(\mathbf{E}_* |p'_r - \mathbf{E}_* p'_r|)^2 \leq \mathbf{E}_* |p'_r - \mathbf{E}_* p'_r|^2 \leq (a_1\gamma_1 t^{1/2})^{-2} a_2(T_t) \leq ct^{-1} a_2 a_1^{-2}, \quad (45)$$

$$|p_r - \mathbf{E}_* p'_r| = p_r |1 - (\gamma_1 t^{1/2})^{-1} \sum_{k \in T_t} k^{-1/2}| \leq ct^{-1}. \quad (46)$$

In (45) we first apply the Cauchy-Schwartz inequality, then use the linearity of variance and the simple inequality $\mathbf{Var} X_k \mathbb{I}_{\{\xi_{3k}=r\}} \leq a_2$. In (46) we use the identity $\mathbf{E}_* X_k \mathbb{I}_{\{\xi_{3k}=r\}} = a_1 p_r$ and (19). From (44), (45), (46) we conclude that $\mathbf{E}_* |p'_r - p_r| = O(t^{-1/2})$. Now (43) implies

$$\mathbf{E}_* |R_3| \mathbb{I}_{\mathcal{D}} \leq \varepsilon + O(|T|t^{-1/2}) = \varepsilon + o(1).$$

□

3.2 Proof of Theorem 2

Here we assume that $\tau(t) := t^\nu$. In the proof below we apply the following simple approximations

$$\begin{aligned} \sum_{k \in T_t} k^{-(1-2\nu)/(2\nu)} &= t^{1/2} \gamma'_1 + rt^{(2\nu)^{-1}-1}, & \sum_{j \in T_k^*} j^{-1/2} &= k^{(2\nu)^{-1}} \gamma'_2 + r' k^{-(2\nu)^{-1}}, \\ \gamma'_1 &:= 2\nu(b^{(2\nu)^{-1}} - a^{(2\nu)^{-1}}), & \gamma'_2 &:= 2(a^{-(2\nu)^{-1}} - b^{-(2\nu)^{-1}}), \end{aligned} \quad (47)$$

where $|r|, |r'| \leq c$. We also make use of relations (20), (22), (23) and (24), which remain valid in the case where $\tau(t) = t^\nu$, and of the inequalities, for $k \in T_t$,

$$|\mathbf{E} b_1(T_k^* \setminus \{t\}) - b_1 \gamma'_2 k^{1/(2\nu)}| \leq cb_1 k^{-1/(2\nu)}, \quad (48)$$

$$\mathbf{E} |b_1(T_k^* \setminus \{t\}) - b_1 \gamma'_2 k^{1/(2\nu)}| \mathbb{I}_{\mathcal{B}_t(\varepsilon)} \leq ck^{1/(2\nu)} (\varepsilon b_1^{1/2} + \mathbf{E} Y_1 \mathbb{I}_{\{Y_1 > \varepsilon^2 t\}}) + cb_1 k^{-1/(2\nu)}. \quad (49)$$

We note that (48) follows from the second identity of (47), and (49) is obtained in the same way as (26) above.

Proof of Theorem 2. Before the proof we introduce some notation. Given $\varepsilon \in (0, 1)$, denote

$$\zeta = \sum_{k \in T_t} \lambda_{kt} \zeta_k, \quad \zeta_k = \beta_k b_1 X_k \mathbb{I}'_k, \quad \mathbb{I}'_k = \mathbb{I}_{\{\beta_k b_1 X_k < \varepsilon\}}, \quad \beta_k = k^{(1-\nu)/(2\nu)} \gamma'_2.$$

Given X, Y , we generate independent Poisson random variables $\eta_k, \hat{\xi}_{3k}$, $k \in T_t$, with (conditional) mean values $\tilde{\mathbf{E}}\eta_k = \lambda_{kt}$, $\tilde{\mathbf{E}}\hat{\xi}_{3k} = \beta_k b_1 X_k$ and independent Bernoulli random variables $\tilde{\mathbb{I}}_k$, $k \in T_t$ with success probabilities

$$\tilde{\mathbf{P}}(\tilde{\mathbb{I}}_k = 1) = 1 - \tilde{\mathbf{P}}(\tilde{\mathbb{I}}_k = 0) = \zeta_k.$$

We assume that, given X, Y , the sequences $\{\mathbb{I}_k, k \in T_t\}$, $\{\tilde{\mathbb{I}}_k, k \in T_t\}$, $\{\eta_k, k \in T_t\}$, $\{\hat{\xi}_{3k}, k \in T_t\}$ are conditionally independent. Next, we introduce random variables

$$\hat{L}_{3t} = \sum_{k \in T_t} \eta_k \hat{\xi}_{3k}, \quad L_{4t} = \sum_{k \in T_t} \mathbb{I}_k \hat{\xi}_{3k}, \quad L_{5t} = \sum_{k \in T_t} \mathbb{I}_k \mathbb{I}'_k \hat{\xi}_{3k}, \quad L_{6t} = \sum_{k \in T_t} \mathbb{I}_k \tilde{\mathbb{I}}_k.$$

Furthermore, we define the random variable L_{7t} as follows. We first generate X, Y . Then, given X, Y , we generate a Poisson random variable with the conditional mean value ζ . The realized value of the Poisson random variable is denoted L_{7t} . Thus, we have $\mathbf{P}(L_{7t} = r) = \mathbf{E}e^{-\zeta} \zeta^r / r!$, for $r = 0, 1, \dots$.

Now we are ready to prove Theorem 2. In the first step of the proof we show that random variables $d(v_t)$ and \hat{L}_{3t} have the same asymptotic distribution (if any). Here we proceed as in the proof of (34), (35) above and make use of (20), (22), (23), (24), (48), (49). In the second step we show that \hat{L}_{3t} converges in distribution to Λ_3 . For this purpose we prove that

$$d_{TV}(\hat{L}_{3t}, L_{4t}) = o(1), \quad \mathbf{E}(L_{4t} - L_{5t}) = o(1), \quad (50)$$

$$d_{TV}(L_{6t}, L_{7t}) = o(1), \quad \mathbf{E}e^{izL_{7t}} - \mathbf{E}e^{iz\Lambda_3} = o(1), \quad (51)$$

for every $-\infty < z < +\infty$, and that there exists $c > 0$, depending only on a, b, ν , such that for any $\varepsilon \in (0, 1)$ we have

$$d_{TV}(L_{5t}, L_{6t}) \leq ca_2 b_1^2 \varepsilon. \quad (52)$$

Let us prove (50), (51), (52). The first bound of (50) is obtained in the same way as the first bound of (34). To show the second bound of (50) we write

$$\tilde{\mathbf{E}}(L_{t5} - L_{t6}) = \sum_{k \in T_t} (1 - \mathbb{I}'_k) \tilde{\mathbf{E}} \mathbb{I}_{kt} \tilde{\mathbf{E}} \hat{\xi}_{3k} = Y_t b_1 t^{-1/2} \sum_{k \in T_t} (1 - \mathbb{I}'_k) X_k^2 \beta_k k^{-1/2}$$

and apply the simple inequality

$$\mathbf{E}X_k^2 (1 - \mathbb{I}'_k) \leq \mathbf{E}X_{\underline{t}}^2 (1 - \mathbb{I}'_{\underline{t}}), \quad k \in T_t. \quad (53)$$

Here we denote $\underline{t} = \min\{k : k \in T_t\}$. We obtain

$$\mathbf{E}(L_{4t} - L_{5t}) = \mathbf{E}\tilde{\mathbf{E}}(L_{4t} - L_{5t}) \leq S_t b_1^2 \mathbf{E}X_{\underline{t}}^2 (1 - \mathbb{I}'_{\underline{t}}) = o(1).$$

Here we denote $S_t = t^{-1/2} \sum_{k \in T_t} \beta_k k^{-1/2}$ and use the simple inequality $S_t \leq c$. Furthermore, we invoke the bound $\mathbf{E}X_{\underline{t}}^2 (1 - \mathbb{I}'_{\underline{t}}) = o(1)$, which holds since $\underline{t} \rightarrow +\infty$ as $t \rightarrow +\infty$

Let us prove (52). Proceeding as in (37), (38) and using the identity $\tilde{\mathbb{I}}_k = \tilde{\mathbb{I}}_k \mathbb{I}'_k$ we write

$$\tilde{d}_{TV}(L_{5t}, L_{6t}) \leq \sum_{k \in T_t} \mathbb{I}'_k \tilde{\mathbf{P}}(\mathbb{I}_k \neq 0) \tilde{d}_{TV}(\hat{\xi}_{3k}, \tilde{\mathbb{I}}_k).$$

Next, we estimate $\mathbb{I}'_k \tilde{d}_{TV}(\hat{\xi}_{3k}, \tilde{\mathbb{I}}_k) \leq \zeta_k^2$, by LeCam's inequality (32), and invoke the inequality $\tilde{\mathbf{P}}(\mathbb{I}_k \neq 0) \leq \lambda_{kt}$. We obtain

$$\tilde{d}_{TV}(L_{5t}, L_{6t}) \leq \sum_{k \in T_t} \mathbb{I}'_k \lambda_{kt} \zeta_k^2 \leq \varepsilon \sum_{k \in T_t} \lambda_{kt} \zeta_k.$$

Here we estimated $\zeta_k^2 \leq \varepsilon \zeta_k$. Now the inequalities

$$d_{TV}(L_{5t}, L_{6t}) \leq \mathbf{E} \tilde{d}_{TV}(L_{5t}, L_{6t}) \leq \varepsilon \sum_{k \in T_t} \mathbf{E} \lambda_{kt} \zeta_k \leq a_2 b_1^2 S_t \varepsilon$$

and $S_t \leq c$ imply (52).

Let us prove the first relation of (51). In view of (17) it suffices to show that $\tilde{d}_{TV}(L_{6t}, L_{7t}) = o_P(1)$. For this purpose we write

$$\tilde{d}_{TV}(L_{6t}, L_{7t}) \leq \mathbb{I}_{\mathcal{A}_1} \tilde{d}_{TV}(L_{6t}, L_{7t}) + \mathbb{I}_{\overline{\mathcal{A}}_1},$$

where $\mathbb{I}_{\overline{\mathcal{A}}_1} = o_P(1)$, see (23), and estimate using LeCam's inequality (32)

$$\mathbb{I}_{\mathcal{A}_1} \tilde{d}_{TV}(L_{6t}, L_{7t}) \leq \mathbb{I}_{\mathcal{A}_1} \sum_{k \in T_t} \tilde{\mathbf{P}}^2(\mathbb{I}_k \tilde{\mathbb{I}}_k = 1) \mathbb{I}'_k \leq Y_t^2 b_1^2 t^{-1} \sum_{k \in T_t} k^{-1} \beta_k^2 X_k^4 = o_P(1)$$

Here we used the simple inequality $t^{-1} \sum_{k \in T_t} k^{-1} \beta_k^2 X_k^4 \leq ct^{-2\nu} \sum_{k \leq bt^\nu} X_k^4$ and the fact that $\mathbf{E} X_1^2 < \infty$ implies the bound $n^{-2} \sum_{k \leq n} X_k^4 = o_P(1)$, as $n \rightarrow +\infty$.

Finally, we show the second relation of (51). We write $\tilde{\mathbf{E}} e^{izL_{7t}} = e^{\zeta(e^{iz}-1)}$ and use the bound

$$Y_t b_1 a_2 \gamma - \zeta = o_P(1). \quad (54)$$

We note that, for any real z , the function $u \rightarrow e^{u(e^{iz}-1)}$ is bounded and uniformly continuous for $u \geq 0$. Therefore, (54) implies the convergence

$$\mathbf{E} e^{izL_{7t}} = \mathbf{E} e^{\zeta(e^{iz}-1)} \rightarrow \mathbf{E} e^{Y_t b_1 a_2 \gamma(e^{iz}-1)} = \mathbf{E} e^{iz\Lambda_3}.$$

It remains to prove (54). We note that (53) implies

$$\zeta = Y_t b_1 \gamma_2'' t^{-1/2} \sum_{k \in T_t} X_k^2 k^{(2\nu)^{-1}-1}. \quad (55)$$

Next, we split $\gamma = \gamma'_1 \gamma''_2$ and invoke the expression for γ'_1 obtained from (47). We obtain

$$Y_t b_1 a_2 \gamma = Y_t b_1 \gamma_2'' t^{-1/2} \sum_{k \in T_t} k^{-(1-2\nu)/(2\nu)} a_2 + o_P(1). \quad (56)$$

We observe that (54) follows from (55), (56) and the bound

$$R := t^{-1/2} \sum_{k \in T_t} (a_2 - X_k^2) k^{(2\nu)^{-1}-1} = o_P(1). \quad (57)$$

In the proof of (57) we use the standard truncation argument. Let $\varepsilon > 0$ and let \hat{R} be defined as R above, but with X_k^2 replaced by $\hat{X}_k^2 = X_k^2 \mathbb{I}_{\{X_k^2 < \varepsilon^2 k\}}$ and a_2 replaced by $\mathbf{E} \hat{X}_k^2$. We have $R = \hat{R} + o_P(1)$ and $\mathbf{P}(\hat{R} > \varepsilon^{1/2}) \leq \varepsilon^{-1} \mathbf{E} \hat{R}^2 \leq c\varepsilon$. Letting $\varepsilon \rightarrow 0$ we obtain $R = o_P(1)$. \square

3.3 Proof of Theorem 3

Before the proof we state an auxiliary lemma.

Lemma 3. Denote $\mathbf{I}_i^x = \mathbb{I}_{\{X_i > i^{1/2}\}}$ and $\mathbf{I}_j^y = \mathbb{I}_{\{Y_j > j^{1/2}\}}$. We have

$$\lambda_{ij}(1 - \mathbf{I}_i^x - \mathbf{I}_j^y) \leq \min\{1, \lambda_{ij}\} \leq \lambda_{ij} \quad (58)$$

Proof of Lemma 3. The inequality $\mathbb{I}_{\{\lambda_{ij} > 1\}} \leq \mathbf{I}_i^x + \mathbf{I}_j^y$ implies

$$\lambda_{ij}(1 - \mathbf{I}_i^x - \mathbf{I}_j^y) \leq \lambda_{ij} - (\lambda_{ij} - 1)\mathbb{I}_{\{\lambda_{ij} > 1\}} = \min\{1, \lambda_{ij}\}.$$

□

Proof of Theorem 3. The proof of (9), (8), (10) is very much the same. Therefore, we only prove (7) and (8).

Before the proof we introduce some notation. Denote

$$T_{st} = T_s \cap T_t, \quad T_{tu} = T_t \cap T_u, \quad T_{stu} = T_s \cap T_t \cap T_u, \quad T = T_s \cup T_t \cup T_u.$$

An attribute w_i is called witness of the edge $v_j \sim v_k$ whenever $\mathbb{I}_{ij}\mathbb{I}_{ik} = 1$. In this case we say that witness w_i realizes the edge $v_j \sim v_k$. Let $\Delta_1 = \{\exists i : \mathbb{I}_{is}\mathbb{I}_{it}\mathbb{I}_{iu} = 1\}$ denote the event that all three edges of the triangle v_s, v_t, v_u are realized by a common witness. Let Δ_2 denote the event that all three edges are realized by different witnesses,

$$\Delta_2 = \{\exists \text{ distinct } i, j, k \text{ such that } \mathbb{I}_{is}\mathbb{I}_{it} = 1, \mathbb{I}_{js}\mathbb{I}_{ju} = 1, \mathbb{I}_{kt}\mathbb{I}_{ku} = 1\}.$$

Let $\Delta = \{v_s \sim v_t, v_s \sim v_u, v_t \sim v_u\}$ denote the event that vertices v_s, v_t, v_u make up a triangle. Introduce events $\mathcal{H}_t = \{v_s \sim v_t, v_t \sim v_u\}$ and $\mathcal{K}_t = \{\exists i \neq j : \mathbb{I}_{it}\mathbb{I}_{is}\mathbb{I}_{jt}\mathbb{I}_{ju} = 1\}$, and random variables

$$\begin{aligned} S &= \sum_{au \leq k \leq bs} \mathbb{I}_{ks}\mathbb{I}_{kt}\mathbb{I}_{ku}, & Q &= \sum_{au \leq i < j \leq bs} \mathbb{I}_{is}\mathbb{I}_{it}\mathbb{I}_{iu}\mathbb{I}_{js}\mathbb{I}_{jt}\mathbb{I}_{ju}, \\ S_t &= \sum_{(i,j) \in I} \mathbb{I}_{it}\mathbb{I}_{is}\mathbb{I}_{jt}\mathbb{I}_{ju}, & Q_t &= \sum_{(i,j) \in I} \sum_{(k,r) \in I, (k,r) \neq (i,j)} \mathbb{I}_{is}\mathbb{I}_{it}\mathbb{I}_{jt}\mathbb{I}_{ju}\mathbb{I}_{kt}\mathbb{I}_{ks}\mathbb{I}_{rt}\mathbb{I}_{ru}. \end{aligned}$$

Here I denote the set of all ordered pairs $(i, j) \in T \times T$ such that $i \neq j$. We remark that every (i, j) indicates a pair (w_i, w_j) of possible witnesses of edges $v_s \sim v_t$ and $v_t \sim v_u$ respectively. We note that for $0 < s < t < u$ satisfying $\lceil au \rceil \leq \lfloor bs \rfloor$ the ratios $t/s, u/t, u/s \in [1, b/a]$. Hence the variables $s, t, u \rightarrow +\infty$ are of the same order of magnitude.

Let us prove (7). We observe that $\Delta_1 \subset \Delta \subset \Delta_1 \cup \Delta_2$. Hence

$$\mathbf{P}(\Delta_1) \leq \mathbf{P}(\Delta) \leq \mathbf{P}(\Delta_1) + \mathbf{P}(\Delta_2). \quad (59)$$

Next, by inclusion exclusion, we write $S - Q \leq \mathbb{I}_{\Delta_1} \leq S$ and estimate

$$\mathbf{E}S - \mathbf{E}Q \leq \mathbf{P}(\Delta_1) \leq \mathbf{E}S. \quad (60)$$

Finally, combining (59) and (60) with the relations

$$\mathbf{E}S = \sum_{au \leq k \leq bs} \frac{\mathbf{E}X_k^3 Y_s Y_t Y_u}{k^{3/2} \sqrt{stu}} + o(t^{-2}) = \frac{a_3 b_1^3}{\sqrt{stu}} \left(\frac{2}{\sqrt{au}} - \frac{2}{\sqrt{bs}} \right) + o(t^{-2}), \quad (61)$$

$$\mathbf{E}Q \leq \sum_{au \leq i < j \leq bs} \mathbf{E}\lambda_{is}\lambda_{it}\lambda_{iu}\lambda_{js}\lambda_{jt}\lambda_{ju} \leq \frac{a_3^2 b_2^3}{stu} \sum_{au \leq i < j \leq bs} \frac{1}{i^{3/2} j^{3/2}} = O(t^{-4}),$$

$$\mathbf{P}(\Delta_2) \leq \mathbf{E} \sum_{i,j,k \in T, i \neq j \neq k} \mathbb{I}_{is}\mathbb{I}_{it}\mathbb{I}_{js}\mathbb{I}_{ju}\mathbb{I}_{kt}\mathbb{I}_{ku} \leq \frac{a_2^3 b_2^3}{stu} \left(\sum_{i \in T} i^{-1} \right)^3 = O(t^{-3}), \quad (62)$$

we obtain asymptotic expression (7) for $p_\Delta = \mathbf{P}(\Delta)$. We note that in the first step of (61) we apply Lemma 3, and in the last step of (62) we use the inequality $\sum_{i \in T} i^{-1} \leq c$.

Let us prove (8). We note that (8) follows from (7) and the relation

$$\mathbf{P}(\mathcal{H}_t) = \mathbf{P}(\Delta) + a_2^2 b_1^2 b_2 \frac{1}{t \sqrt{su}} \delta_{t|su} + o(t^{-2}). \quad (63)$$

It remains to show (63). From the identity $\mathcal{H}_t = \Delta_1 \cup \mathcal{K}_t$ we obtain

$$\mathbf{P}(\mathcal{H}_t) = \mathbf{P}(\Delta_1) + \mathbf{P}(\mathcal{K}_t) - \mathbf{P}(\Delta_1 \cap \mathcal{K}_t). \quad (64)$$

Next, by inclusion exclusion, we write $S_t - Q_t \leq \mathbb{I}_{\mathcal{K}_t} \leq S_t$. These inequalities imply

$$\mathbf{E}S_t - \mathbf{E}S_t(1 - \mathbb{I}_{\mathcal{D}_\varepsilon}) - \mathbf{E}Q_t \mathbb{I}_{\mathcal{D}_\varepsilon} \leq \mathbf{E}\mathbb{I}_{\mathcal{K}_t} \mathbb{I}_{\mathcal{D}_\varepsilon} \leq \mathbf{P}(\mathcal{K}_t) \leq \mathbf{E}S_t. \quad (65)$$

Here the event $\mathcal{D}_\varepsilon = \{Y_t \leq \varepsilon t\}$ and $\varepsilon \in (0, 1)$ is non-random. In the remaining part of the proof we show that

$$\mathbf{E}S_t = a_2^2 b_1^2 b_2 \frac{1}{t \sqrt{su}} \delta_{t|s,u} + o(t^{-2}), \quad (66)$$

$$\mathbf{P}(\Delta_1 \cap \mathcal{K}_t) = O(t^{-3}), \quad (67)$$

and that there exists $c^* > 0$ which does not depend on s, t, u and ε such that, for any $\varepsilon \in (0, 1)$,

$$\mathbf{E}Q_t \mathbb{I}_{\mathcal{D}_\varepsilon} \leq c^* \varepsilon t^{-2} + O(t^{-3}), \quad \mathbf{E}S_t(1 - \mathbb{I}_{\mathcal{D}_\varepsilon}) = o(t^{-2}). \quad (68)$$

We observe that (63) follows from (64), (65), (66) and the bounds (67), (68).

Let us prove (66). Since the product $\bar{p}_{ij} := p_{is}p_{it}p_{jt}p_{ju}$ is non zero whenever $i \in T_{st}$ and $j \in T_{tu}$, we have

$$\mathbf{E}S_t = \mathbf{E} \sum_{(i,j) \in I} \bar{p}_{ij} = \mathbf{E} \sum_{(i,j): i \in T_{st}, j \in T_{tu}, i \neq j} \bar{p}_{ij}. \quad (69)$$

It is convenient to split the set $\{(i, j) : i \in T_{st}, j \in T_{tu}, i \neq j\} = \mathbb{T}_1 \cup \dots \cup \mathbb{T}_4$ where

$$\begin{aligned} \mathbb{T}_1 &= (T_{st} \setminus T_{tu}) \times T_{tu}, & \mathbb{T}_2 &= T_{stu} \times (T_{tu} \setminus T_s), \\ \mathbb{T}_3 &= \{(i, j) : i, j \in T_{stu}, i < j\}, & \mathbb{T}_4 &= \{(i, j) : i, j \in T_{stu}, j < i\}. \end{aligned}$$

and write sum (69) in the form

$$\tilde{\mathbf{E}}S_t = S_{t1} + \dots + S_{t4}, \quad S_{tk} := \sum_{(i,j) \in \mathbb{T}_k} \bar{p}_{ij}. \quad (70)$$

Now (66) follows from (70) and the relations, for $1 \leq k \leq 4$,

$$\begin{aligned} \mathbf{E}S_{tk} &= \mathbf{E} \sum_{(i,j) \in \mathbb{T}_k} \lambda_{is}\lambda_{it}\lambda_{jt}\lambda_{ju} + o(t^{-2}) = a_2^2 b_1^2 b_2 \frac{1}{t\sqrt{su}} \sum_{(i,j) \in \mathbb{T}_k} \frac{1}{ij} + o(t^{-2}), \\ &\quad \sum_{1 \leq k \leq 4} \sum_{(i,j) \in \mathbb{T}_k} \frac{1}{ij} = \delta_{t|su} + O(t^{-1}). \end{aligned} \quad (71)$$

In the first step of (71) we used Lemma 3.

Let us prove the first bound of (68). We split the collection of vectors (i, j, k, r)

$$\mathbb{Q} = \{(i, j, k, r) \in T^4 : \text{such that } i \neq j, k \neq r \text{ and } (i, j) \neq (k, r)\}$$

into five non intersecting pieces $\mathbb{Q} = \mathbb{Q}_1 \cup \dots \cup \mathbb{Q}_5$, where

$$\begin{aligned} \mathbb{Q}_1 &= \{(i, j, k, r) : i = k\} \cap \mathbb{Q}, & \mathbb{Q}_2 &= \{(i, j, k, r) : i = r\} \cap \mathbb{Q}, \\ \mathbb{Q}_3 &= \{(i, j, k, r) : j = k\} \cap \mathbb{Q}, & \mathbb{Q}_4 &= \{(i, j, k, r) : j = r\} \cap \mathbb{Q}, \end{aligned}$$

and $\mathbb{Q}_5 = \{(i, j, k, r) : \text{all } i, j, k, r \text{ are distinct}\} \cap \mathbb{Q}$, and write

$$Q_t = \sum_{1 \leq z \leq 5} Q_{tz}, \quad Q_{tz} = \sum_{(i,j,k,r) \in \mathbb{Q}_z} \mathbb{I}_{is}\mathbb{I}_{it}\mathbb{I}_{jt}\mathbb{I}_{ju}\mathbb{I}_{ks}\mathbb{I}_{kt}\mathbb{I}_{rt}\mathbb{I}_{ru}.$$

Denote $\tilde{\mathbb{Q}} = \{(i, j, r) \in T^3 : \text{all } i, j, r \text{ are distinct}\}$. Observing that the typical summand of the sum Q_{t1} is $\mathbb{I}_{is}\mathbb{I}_{it}\mathbb{I}_{jt}\mathbb{I}_{ju}\mathbb{I}_{rt}\mathbb{I}_{ru}$ (since $i = k$), we write

$$\begin{aligned} \mathbf{E}Q_{t1}\mathbb{I}_{\mathcal{D}_\varepsilon} &\leq \mathbf{E} \sum_{(i,j,r) \in \tilde{\mathbb{Q}}} \lambda_{is}\lambda_{it}\lambda_{jt}\lambda_{ju}\lambda_{rt}\lambda_{ru}\mathbb{I}_{\mathcal{D}_\varepsilon} \\ &\leq \frac{a_2^3}{s^{1/2}t^{3/2}u} \mathbf{E}Y_s Y_t^3 Y_u^2 \mathbb{I}_{\mathcal{D}_\varepsilon} \left(\sum_{i \in T} \frac{1}{i}\right)^3 \\ &\leq c^3 \varepsilon \frac{a_2^3}{s^{1/2}t^{1/2}u} \mathbf{E}Y_s Y_t^2 Y_u^2 \\ &\leq c' \varepsilon t^{-2}. \end{aligned}$$

Here used inequalities $Y_t t^{-1} \mathbb{I}_{\mathcal{D}_\varepsilon} \leq \varepsilon$ and $\sum_{i \in T} \frac{1}{i} \leq c$. Similarly, we prove the inequality $\mathbf{E}Q_{t4}\mathbb{I}_{\mathcal{D}_\varepsilon} \leq c' \varepsilon t^{-2}$. Furthermore, observing that the typical summand of the sum Q_{t2} is $\mathbb{I}_{is}\mathbb{I}_{it}\mathbb{I}_{iu}\mathbb{I}_{jt}\mathbb{I}_{ju}\mathbb{I}_{ks}\mathbb{I}_{kt}$ (since $i = r$), we write

$$\begin{aligned} \mathbf{E}Q_{t2}\mathbb{I}_{\mathcal{D}_\varepsilon} &\leq \mathbf{E} \sum_{(i,j,k) \in \tilde{\mathbb{Q}}} \lambda_{is}\lambda_{it}\lambda_{iu}\lambda_{jt}\lambda_{ju}\lambda_{ks}\lambda_{kt}\mathbb{I}_{\mathcal{D}_\varepsilon} \\ &\leq \frac{a_3 a_2^2}{s t^{3/2} u} \mathbf{E}Y_s^2 Y_t^3 Y_u^2 \mathbb{I}_{\mathcal{D}_\varepsilon} \left(\sum_{i \in T} \frac{1}{i}\right)^2 \left(\sum_{i \in T} \frac{1}{i^{3/2}}\right) \\ &\leq c^3 \frac{a_3 a_2^2}{stu} \mathbf{E}Y_s^2 Y_t^2 Y_u^2. \end{aligned}$$

In the last step we used inequalities $Y_t t^{-1} \mathbb{I}_{\mathcal{D}_\varepsilon} \leq 1$ and $\sum_{i \in T} \frac{1}{i^{3/2}} \leq ct^{-1/2}$. Hence, $\mathbf{E}Q_{t2}\mathbb{I}_{\mathcal{D}_\varepsilon} =$

$O(t^{-3})$. Similarly, we prove the bound $\mathbf{E}Q_{t3}\mathbb{I}_{\mathcal{D}_\varepsilon} = O(t^{-3})$. Finally, we estimate

$$\begin{aligned}\mathbf{E}Q_{t5}\mathbb{I}_{\mathcal{D}_\varepsilon} &\leq \mathbf{E} \sum_{(i,j,k,r) \in \mathbb{Q}_5} \lambda_{is}\lambda_{is}\lambda_{jt}\lambda_{ju}\lambda_{ks}\lambda_{kt}\lambda_{rt}\lambda_{ru}\mathbb{I}_{\mathcal{D}_\varepsilon} \\ &\leq \frac{a_2^4}{st^2u} \mathbf{E}Y_s^2Y_t^4Y_u^2\mathbb{I}_{\mathcal{D}_\varepsilon} \left(\sum_{i \in T} \frac{1}{i}\right)^4 \\ &\leq c'\varepsilon^2 t^{-2}.\end{aligned}$$

In the last step we used the inequality $Y_t^2t^{-2}\mathbb{I}_{\mathcal{D}_\varepsilon} \leq \varepsilon^2$.

Let us prove the second bound of (68). We have

$$\mathbf{E}S_t(1 - \mathbb{I}_{\mathcal{D}_\varepsilon}) \leq \mathbf{E} \sum_{i,j \in T, i \neq j} \lambda_{is}\lambda_{it}\lambda_{jt}\lambda_{ju}(1 - \mathbb{I}_{\mathcal{D}_\varepsilon}) \leq \frac{a_2^2 b_1^2}{st} \mathbf{E}Y_t^2 \mathbb{I}_{\{Y_t \geq \varepsilon t\}} \left(\sum_{i \in T} i^{-1}\right)^2 = o(t^{-2}).$$

Let us prove (67). The inequalities $\mathbb{I}_{\mathcal{K}_t} \leq S_t$, $\mathbb{I}_{\Delta_1} \leq S$ and $S \leq \tilde{S}$, where

$$\tilde{S} = \sum_{k \in T} \mathbb{I}_k^* \quad \text{and} \quad \mathbb{I}_k^* = \mathbb{I}_{ks}\mathbb{I}_{kt}\mathbb{I}_{ku},$$

imply $\mathbf{P}(\Delta_1 \cap \mathcal{K}_t) = \mathbf{E}\mathbb{I}_{\Delta_1}\mathbb{I}_{\mathcal{K}_t} \leq \mathbf{E}S_t\tilde{S}$. We show that $\mathbf{E}S_t\tilde{S} = O(t^{-3})$. We split $S_t\tilde{S} = \tilde{S}_1 + \tilde{S}_2$,

$$\tilde{S}_1 = \sum_{i \in T} \sum_{j \in T \setminus \{i\}} \mathbb{I}_{is}\mathbb{I}_{it}\mathbb{I}_{jt}\mathbb{I}_{ju}(\mathbb{I}_i^* + \mathbb{I}_j^*), \quad \tilde{S}_2 = \sum_{(i,j,k) \in \tilde{\mathbb{Q}}} \mathbb{I}_{is}\mathbb{I}_{it}\mathbb{I}_{jt}\mathbb{I}_{ju}\mathbb{I}_k^*,$$

and estimate

$$\begin{aligned}\mathbf{E}\tilde{S}_1 &\leq \mathbf{E} \sum_{i \in T} \sum_{j \in T \setminus \{i\}} \lambda_{is}\lambda_{it}\lambda_{jt}\lambda_{ju}(\lambda_{iu} + \lambda_{js}) = O(t^{-3}), \\ \mathbf{E}\tilde{S}_2 &\leq \mathbf{E}\tilde{S}'_2 \leq \mathbf{E} \sum_{(i,j,k) \in \tilde{\mathbb{Q}}} \lambda_{is}\lambda_{it}\lambda_{jt}\lambda_{ju}\lambda_{ks}\lambda_{ku} = O(t^{-3}).\end{aligned}\tag{72}$$

Here \tilde{S}'_2 is defined in the same way as \tilde{S}_2 , but with \mathbb{I}_k^* replaced by $\mathbb{I}'_k = \mathbb{I}_{ks}\mathbb{I}_{ku}$. □

3.4 Proof of (11)

We only give a sketch of the proof. Let $s < t$ satisfy the inequality $\lceil at \rceil \leq \lceil bs \rceil$. An attribute w_i is called witness of the edge $v_s \sim v_t$ whenever $\mathbb{I}_{it}\mathbb{I}_{is} = 1$. The sums

$$e_{st} = \sum_{i \in T_s \cap T_t} \mathbb{I}_{is}\mathbb{I}_{it} \quad \text{and} \quad q_{st} = \sum_{\{i,j\} \subset T_s \cap T_t} \mathbb{I}_{is}\mathbb{I}_{it}\mathbb{I}_{js}\mathbb{I}_{jt}$$

count witnesses and pairs of witnesses of the edge $v_s \sim v_t$, respectively. We write, by inclusion-exclusion,

$$e_{st} - q_{st} \leq \mathbb{I}_{\{v_s \sim v_t\}} \leq e_{st} \tag{73}$$

and note that the quadratic term q_{st} is negligibly small. Hence, we approximate

$$\mathbb{I}_{\{v_s \sim v_t\}} = e_{st}(1 + o_P(1)), \quad \mathbf{P}(v_s \sim v_t) = (1 + o(1))\mathbf{E}e_{st}. \tag{74}$$

Given t and $i, j \in T_t$, we denote $T_{it}^* = T_i^* \setminus \{t\}$ and introduce random variables

$$u_{it} = \sum_{k \in T_{it}^*} \mathbb{I}_{ik}, \quad z_{ijt} = \sum_{k \in T_{it}^* \cap T_{jt}^*} \mathbb{I}_{ik} \mathbb{I}_{jk}, \quad L_t = \sum_{i \in T_t} \mathbb{I}_{it} u_{it}, \quad Q_t = \sum_{\{i,j\} \subset T_t} \mathbb{I}_{it} \mathbb{I}_{jt} z_{ijt}.$$

We remark that L_t counts pairs $(v_s \sim v_t; w_i)$, where w_i is a witness of the edge $v_s \sim v_t$ in $G_{X,Y}$, for some $v_s \in W \setminus \{v_t\}$. In particular, we have $d(v_t) \leq L_t$. Similarly, Q_t counts all triples $(v_s \sim v_t; w_i, w_j)$, where w_i and w_j are distinct witnesses of an edge $v_s \sim v_t$. Note that a neighbour v_s of v_t , which has k witnesses of the edge $\{v_s \sim v_t\}$, contributes 1 to the number $d(v_t)$ of neighbours of v_t . It contributes k to the sum L_t and it contributes $\binom{k}{2}$ to the sum Q_t . Hence, we always have

$$L_t - Q_t \leq d(v_t) \leq L_t.$$

We note that the quadratic term Q_t is negligibly small and approximate $d(v_t) = L_t(1 + o_P(1))$. Combining this approximation with (74) we obtain, for $r = 1, 2$ and $u = s, t$,

$$\mathbf{E}_{st} d^r(v_u) = (\mathbf{E} e_{st})^{-1} \mathbf{E} e_{st} L_u^r + o(1) \quad \text{and} \quad \mathbf{E}_{st} d(v_s) d(v_t) = (\mathbf{E} e_{st})^{-1} \mathbf{E} e_{st} L_s L_t + o(1). \quad (75)$$

Next we evaluate expectations in the right-hand sides of (75). A straightforward but tedious calculation shows that

$$\begin{aligned} \mathbf{E} e_{st} &= \Theta(1 + o(1)) h_1, \\ \mathbf{E} e_{st} L_s &= \mathbf{E} e_{st} L_t + o(\Theta) = \Theta(1 + o(1))(h_1 + h_2 + 2h_3), \\ \mathbf{E} e_{st} L_s^2 &= \mathbf{E} e_{st} L_t^2 + o(\Theta) = \Theta(1 + o(1))(h_1 + 3h_2 + 6h_3 + h_4 + 6h_5 + 4h_6), \\ \mathbf{E} e_{st} L_s L_t &= \Theta(1 + o(1))(h_1 + 3h_2 + 4h_3 + h_4 + 4h_5 + 4h_7). \end{aligned}$$

Here we denote $\Theta = (st)^{-1/2} \ln(bs/at)$. We recall that h_i are defined in (12) above. Now (11) follows from (75).

Acknowledgement. Research of M. Bloznelis was supported by the Research Council of Lithuania grant MIP-053/2011.

References

- [1] M. Bloznelis, Degree distribution of a typical vertex in a general random intersection graph, Lithuanian Mathematical Journal 48 (2008), 38–45.
- [2] M. Bloznelis, Degree and clustering coefficient in sparse random intersection graphs, to appear in The Annals of Applied Probability.
- [3] M. Bloznelis, V. Kurauskas, Clustering function: a measure of social influence, <http://arxiv.org/abs/1207.4941>
- [4] T. Britton, M. Deijfen, M. Lindholm, and N. A. Leras, Epidemics on random graphs with tunable clustering. J. Appl. Prob. 45 (2008), 743–756.
- [5] T. Britton, M. Lindholm, Dynamic random networks in dynamic populations, J. Statistical Physics 139 (2010), 518–535.
- [6] T. Britton, M. Lindholm, and T. S. Turova, A dynamic random network in a dynamic population: asymptotic properties, J. Appl. Probab. 48 (2011), 1163–1178.

- [7] M. Deijfen and W. Kets, Random intersection graphs with tunable degree distribution and clustering, *Probab. Engrg. Inform. Sci.* 23 (2009), 661–674.
- [8] L. Eschenauer and V. D. Gligor, A key-management scheme for distributed sensor networks, in: *Proceedings of the 9th ACM Conference on Computer and Communications Security* (2002), 41–47.
- [9] E. Godehardt and J. Jaworski, Two models of random intersection graphs and their applications, *Electronic Notes in Discrete Mathematics* 10 (2001), 129–132.
- [10] E. Godehardt, J. Jaworski, and K. Rybarczyk, Clustering coefficients of random intersection graphs, in: *Studies in Classification, Data Analysis and Knowledge Organization*, Springer, Berlin–Heidelberg–New York, 2011. To appear.
- [11] J. Jaworski, M. Karoński, and D. Stark, The degree of a typical vertex in generalized random intersection graph models, *Discrete Mathematics* 306 (2006), 2152–2165.
- [12] J. Jaworski and D. Stark, The vertex degree distribution of passive random intersection graph models, *Combinatorics, Probability and Computing* 17 (2008), 549–558.
- [13] M. Karoński, E. R. Scheinerman, and K. B. Singer-Cohen, On random intersection graphs: The subgraph problem, *Combinatorics, Probability and Computing* 8 (1999), 131–159.
- [14] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, Random graph models of social networks, *Proc. Natl. Acad. Sci. USA*, 99 (Suppl. 1) (2002), 2566–2572.
- [15] K. B. Singer-Cohen, Random intersection graphs, PhD thesis, Department of Mathematical Sciences, The Johns Hopkins University, 1995
- [16] D. Stark, The vertex degree distribution of random intersection graphs, *Random Structures and Algorithms* 24 (2004), 249–258.
- [17] J. M. Steele, Le Cam’s inequality and Poisson approximations, *The American Mathematical Monthly* 101 (1994), 48–54.
- [18] T.S. Turova, Dynamical random graphs with memory, *Physical Review E*, 65 (2002), 066102.
- [19] T.S. Turova, Phase transitions in Dynamical Random Graphs, *J. Statistical Physics*, 123 (2007), 1007–1032.